

WHAT IS CLAIMED IS:

1. A method of deriving sequence annotations for sequences in a genomics or proteomics database, said method comprising:

5 modeling the three dimensional structure of at least one protein encoded by a sequence in said database;
modeling an interaction between at least one ligand and said modeled three dimensional structure; and
deriving an annotation from calculated characteristics of said interaction.

10 2. The method of Claim 1, wherein said sequences comprise nucleic acid sequences.

3. The method of Claim 1, wherein said sequences comprise amino acid sequences.

15 4. A method of annotating sequences in a genomics or proteomics database, said method comprising:

selecting a set of sequences from said database;
obtaining a structural model of each protein encoded by the set of sequences;
selecting a set of ligand molecules;
separately modeling an interaction between each ligand and each structural protein model;
20 deriving a value indicative of the strength of interaction between each ligand molecule and each protein model; and
storing the values in association with the sequences in the database.

25 5. The method of Claim 4, wherein said value is binary.

6. The method of Claim 4, wherein the ligand molecules making up said set are chemically diverse.

7. A method of making a functional association between first and second protein molecules, said method comprising:

30 retrieving a first series of values representative of binding strength between said first protein and a set of ligand molecules;

retrieving a second series of values representative of binding strength between said second protein and said set of ligand molecules;

comparing said first series of values with said second series of values.

8. A computer readable medium storing a plurality of gene sequences, at least a first one of which has one or more annotations stored in association therewith, wherein said annotations comprise a set of values indicative of the predicted strength of binding between a protein encoded by said first gene sequence and a corresponding set of chemically diverse ligand molecules.

9. A method of characterizing a protein, said method comprising:
modeling an interaction between said protein and a ligand molecule;
deriving a value indicative of binding strength between said protein and said ligand molecule;
repeating said modeling and deriving for one or more additional ligand molecules; and

storing said values as an associated set so as to form an interaction fingerprint characterizing chemical behavior of said protein.

10. A method of comparing first and second protein molecules, said method comprising:

retrieving a first set of values representative of binding strength between said first protein and a corresponding set of ligands;

retrieving a second set of values representative of binding strength between said second protein and said set of ligands; and

comparing said first set of values to said second set of values.

11. The method of Claim 10, wherein said values are binary indications of the either the presence of binding or the absence of binding.

12. The method of Claim 10, wherein said comparing comprises multiplying values in each set corresponding to the same ligand.

13. A method of identifying a target protein for pharmaceutical intervention comprising:

(a) selecting a first potential target protein;

(b) retrieving a first interaction fingerprint comprising a set of values representative of binding strength between said potential target protein and a corresponding set of ligands;

(c) retrieving a different interaction fingerprint comprising a set of values representative of binding strength between a different protein and said set of ligands;

(d) comparing said first interaction fingerprint with said second interaction fingerprint;

(e) repeating steps (c) and (d) for a plurality of different proteins encoded by a selected genome.

14. The method of Claim 13, wherein steps (c) and (d) are repeated for substantially all proteins encoded by said selected genome.

15. The method of Claim 14, wherein said selected genome is the human genome.

16. A system for biological research comprising:

a database storing both gene sequences and interaction fingerprints characterizing chemical behavior of at least some proteins encoded by said gene sequences; and

a search and computation engine configured to retrieve and compare said interaction fingerprints.

17. A method of assessing ligand interactions, said method comprising:

selecting a ligand; and

modeling the interaction of the ligand with a plurality of protein models spanning substantially an entire genome.

18. The method of Claim 17, further comprising:

selecting n additional ligands, where n is an integer of one or more; and

modeling the interaction of the n additional ligands with said plurality of protein models spanning substantially an entire genome.

19. The method of Claim 17, wherein said ligand is a drug candidate that is tested for toxic response by assessing the modeled protein/ligand interactions.